

Exploratory Data Analysis on YouTube data.

Domain: Social media

Context and Content: In a fairly recent move by Youtube, it announced the decision to hide the number of dislikes from users around November 2021. However, the official YouTube Data API allowed you to get information about dislikes until December 13, 2021. Doing an EDA-exercise can help to draw some unseen insights from this dataset.

SUBMITTED BY:

NITHISHWAR

1. Import required libraries and read the provided dataset (youtube_dislike_dataset.csv) and retrieve top 5 and bottom 5 records.

In []: 1 import pandas as pd

In [3]: 1 # -- Reading the csv file
2
3 df = pd.read_csv(r"C:\Users\Withish\Desktop\GL\Python\week-3\Graded Project week 3\youtube_dislike_dataset.csv")
4 df

Out[3]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuW	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football socc alemn Bun se
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGleIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Ava Quality Music/Moto
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRCm4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1*YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxyvulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101 JO1 TheSTAF
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjIQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining w Jersey highwa C
...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Calling My Calling h
37418	zziBybeSAtw	PELICANS at LAKERS FULL GAME HIGHLIGHTS Ja...	UCWJ2IWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G Basketball 0022000187 L
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxIP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAMOO WAW MAM WAW Where
37420	zzmQEb0Em5l	FELIPE ESCUDEIRO- Master Podcast #12	UC8NjnNWMsRqq11NyVHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	masterpodc lord vinheteir

```
In [4]: 1 df1 = df.copy() # -- For future use -- Backup Dataframe
        2 df1
```

Out[4]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football socc alemn Bun se
1	--14w5SOEU	Migos - Avalanche (Official Video)	UCGIeIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avi Quality Music/Moto
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEDjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyulMA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101 JO1 TheSTAF
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQldVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining w Jersey highwa C
...	
37417	zzd4ydaFGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Calling My Calling M
37418	zziBybeSAtw	PELICANS at LAKERS FULL GAME HIGHLIGHTS Ja...	UCWJ2IWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G Basketball 0022000187 L
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxiP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAMOO WAW MAM WAW Where
37420	zzmQEb0Em5I	FELLIPE ESCUDEIRO- Master Podcast #12	UC8NjnNWMsRqq11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	masterpodc lord vinheteir
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	Spurs Tot Hotspur Tot Leic

37422 rows × 12 columns

In [5]: 1 df.head() # Top 5 records

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tags
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football soccer f1bol alemn Bundesliga season ...
1	-14w5SOEUu	Migos - Avalanche (Official Video)	UCG1eIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avalanche Quality Control Music/Motown R...
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxyvulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101JAPAN JO1 TheSTAR STA...
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjIQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining wall New Jersey highway Direct Conne...

In [6]: 1 df.tail() # -- Bottom 5 records

37418	zziBybeSATw	LAKERS FULL GAME HIGHLIGHTS Ja...	UCWJ2iWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	League Basketbal game- 0022000187 Lakers...
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxIP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAMOC WAW WAW MAMAMOC WAW Where Are We Now...
37420	zzmQEb0Em5l	FELLIPE ESCUDERO- Master Podcast #12	UC8NjnNWMsRqQ11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	master masterpodcas lord lord vinheteiro z z ...
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	Spurs Tottenham Hotspur Tottenham Leicester ...

2. Check the info of the dataframe and write your inferences on data types and shape of the dataset.

```
In [7]: 1 # -- To pull all the information about the dataframe
        2
        3 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37422 entries, 0 to 37421
Data columns (total 12 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   video_id        37422 non-null  object
1   title           37422 non-null  object
2   channel_id      37422 non-null  object
3   channel_title   37422 non-null  object
4   published_at    37422 non-null  object
5   view_count      37422 non-null  int64
6   likes           37422 non-null  int64
7   dislikes        37422 non-null  int64
8   comment_count   37422 non-null  int64
9   tags            37422 non-null  object
10  description      37422 non-null  object
11  comments        37264 non-null  object
dtypes: int64(4), object(8)
memory usage: 3.4+ MB
```

```
In [8]: 1 df.dtypes # -- This Data frame has 8 object columns and 4 int columns
```

```
Out[8]: video_id      object
        title         object
        channel_id    object
        channel_title  object
        published_at   object
        view_count     int64
        likes          int64
        dislikes       int64
        comment_count  int64
        tags           object
        description    object
        comments       object
        dtype: object
```

```
In [9]: 1 df.shape    # --This dataframe has 37422 rows and 12 columns
```

```
Out[9]: (37422, 12)
```

Inference

```
In [10]: 1 '''
2 So,This Data frame has 8 object columns and 4 int columns according to df.dtypes
3
4 Then,This dataframe has 37422 rows and 12 columns according to df.shape
5
6 '''
```

```
Out[10]: '\nSo,This Data frame has 8 object columns and 4 int columns according to df.dtypes\n\nThen,This dataframe has 37422 rows and 1
2 columns according to df.shape\n\n'
```

3. Check for the Percentage of the missing values and drop or impute them.

```
In [11]: 1 # --To check that the missing values are present or not in the data frame
2
3 df.isnull()
```

```
Out[11]:
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tags	description	comments
0	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False
...
37417	False	False	False	False	False	False	False	False	False	False	False	False
37418	False	False	False	False	False	False	False	False	False	False	False	False
37419	False	False	False	False	False	False	False	False	False	False	False	False
37420	False	False	False	False	False	False	False	False	False	False	False	False
37421	False	False	False	False	False	False	False	False	False	False	False	False

37422 rows × 12 columns

```
In [12]: 1 # --To count the missing values
```

```
2  
3 df.isnull().sum()
```

```
Out[12]: video_id      0  
title      0  
channel_id  0  
channel_title  0  
published_at  0  
view_count  0  
likes      0  
dislikes   0  
comment_count  0  
tags      0  
description 0  
comments   158  
dtype: int64
```

```
In [13]: 1 # --To count the missing values in the 'comments' column
```

```
2  
3 df['comments'].isnull().sum()
```

```
Out[13]: 158
```

```
In [14]: 1 # -- To count the unique values
```

```
2  
3 df['comments'].value_counts()
```

de la suite... Enfin une course de bonhomme, j'espère que ça va continuer F.Montagny et J.Febreau juste incroyable beaucoup mieux qu'avec J.Villeneuve Le départ a été incroyable parce que Hamilton était obligé de passer Verstappen au 1er tour pour avoir une chance de gagner, vu le rythme de la Red Bull. Vous connaissez les risques de ce sport, dommage pour Max bon rétablissement mais nous on veut de la bataille, du spectacle. C'est d'ailleurs pour ça que vous et moi sommes là à regarder. Qu'on l'aime ou pas Respect Hamilton aussi il a tenu et malgré tout il a su rebondir. Il a travaillé toute la semaine pour ce GP. Ne lui en voulez pas de se battre Toto Wolff à la fin de ce GP : \nC'est la faute Valtteri Certes le départ était tonitruant et à bout de souffle et Hamilton a gagné sur ses terres bravo lui même si Leclerc méritait lui aussi la victoire tout comme Verstappen qui était le plus rapide mais ce n'est pas la même course et niveau ambiance les tifosi sont imbattables Forza Italia et Ferrari sempre Quelle course de Charles tellement dommage pour la victoire On va pas se mentir, malgré le fait que ça a fini en crash, le début de ce GP était titanesque.

1
3K Comments Aur 30k likes ka target pura krdo ???? Chahiye part 2 ???? So funny prince bhai keep it up WE NEED PART 2 PLEASE We need part 2 Amazing video Beautiful video WE NEED Part 2 We need part 2 We need to Part 2 We need part 2 please We need part 2 plz This is a very funny video I liked it and Imagine getting heart prince pathania and yash srivastav We need part 2 We need part 2 we need part 2 We need part 3 We need part 2 Mjaa aaya Cycle mein bhoot part 2 kab banega Prince bhai We need part 2

1
Can we take a moment to appreciate the woman who with the real and useful hack, and that Ann has asked us to support her? Truly wonderful. Can we just appreciate that she made her book in both American and non American measurements?!\nLike the amount of work that went into this book is incredible. I hope that the book is a success and that it helps to support the work of the woman who has asked us to support her?

```
In [15]: 1 # -- Checking for the percentage
2
3 a = (df.isnull().sum()/len(df)) * 100
4 a
5
6 # -- So it is
```

```
Out[15]: video_id      0.000000
title      0.000000
channel_id  0.000000
channel_title  0.000000
published_at  0.000000
view_count  0.000000
likes      0.000000
dislikes    0.000000
comment_count  0.000000
tags      0.000000
description  0.000000
comments    0.422212
dtype: float64
```

```
In [16]: 1 round(1/12) # -- comments column has 4% of nul values we can drop the 'comments' column as it will not affect the datafra
2
```

```
Out[16]: 0
```

```
In [17]: 1 df.drop(columns='comments') # -- Droping 'comments' column
```

```
Out[17]:
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football socc alemn Bun se
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIeIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Av Quality Music/Moto
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101 JO1 TheSTAF
4	--DKkzVWvH-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjIQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining w Jersey highwa C
...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Calling My Calling h

4. Check the statistical summary of both numerical and categorical columns and write your inferences.

```
In [18]: 1 numcol = df.select_dtypes(exclude='O') # --Containing only numeric values
          2
          3 catcol = df.select_dtypes(include='O') # --Containing only catagorical columns
```

```
In [19]: 1 # -- Statistical summary of numerical column
          2
          3 numcol.describe()
```

```
Out[19]:
```

	view_count	likes	dislikes	comment_count
count	3.742200e+04	3.742200e+04	3.742200e+04	3.742200e+04
mean	5.697838e+06	1.668147e+05	4.989862e+03	9.924930e+03
std	2.426622e+07	5.375670e+05	3.070824e+04	1.171003e+05
min	2.036800e+04	0.000000e+00	0.000000e+00	0.000000e+00
25%	5.122970e+05	1.323350e+04	2.810000e+02	9.000000e+02
50%	1.319078e+06	4.233050e+04	7.960000e+02	2.328000e+03
75%	3.670231e+06	1.304698e+05	2.461750e+03	6.184000e+03
max	1.322797e+09	3.183768e+07	2.397733e+06	1.607103e+07

```
In [20]: 1 numcol.skew()
```

```
Out[20]: view_count      20.327887
likes      16.214994
dislikes   31.939953
comment_count      89.049536
dtype: float64
```

```
In [21]: 1 numcol.kurt()
```

```
Out[21]: view_count      705.914574
likes      543.374196
dislikes   1709.867836
comment_count      10587.249503
dtype: float64
```

```
In [22]: 1 numcol.corr()
```

```
Out[22]:
```

	view_count	likes	dislikes	comment_count
view_count	1.000000	0.784229	0.684469	0.464392
likes	0.784229	1.000000	0.679817	0.580953
dislikes	0.684469	0.679817	1.000000	0.391739
comment_count	0.464392	0.580953	0.391739	1.000000

```
In [23]: 1 numcol.median()
```

```
Out[23]: view_count      1319078.5  
likes          42330.5  
dislikes         796.0  
comment_count   2328.0  
dtype: float64
```

```
In [24]: 1 numcol.mode()
```

```
Out[24]:
```

	view_count	likes	dislikes	comment_count
0	784442.0	2020	88.0	0.0
1	NaN	3504	NaN	NaN
2	NaN	4691	NaN	NaN

```
In [25]: 1 numcol.var()
```

```
Out[25]: view_count      5.888493e+14  
likes          2.889783e+11  
dislikes         9.429959e+08  
comment_count   1.371247e+10  
dtype: float64
```

```
In [26]: 1 # -- Statistical summary of categorical column  
2  
3 catcol.describe()
```

```
Out[26]:
```

	video_id	title	channel_id	channel_title	published_at	tags	description	comments
count	37422	37422	37422	37422	37422	37422	37422	37264
unique	37422	37113	10961	10883	36772	28799	35630	37264
top	--0bCF-iK2E	www	UCNAf1k0yljyGu3k9BwAg3lg	Sky Sports Football	2020-10-16 04:00:10		Respect to Dortmund fans,must be sad losing hi...	
freq	1	21	533	533	6	3817	589	1

5. Convert datatype of column published_at from object to pandas datetime.

```
In [29]: 1 df['published_at']
```

```
Out[29]: 0      2021-07-01 10:00:00
1      2021-06-10 16:00:00
2      2021-09-20 01:03:32
3      2021-03-03 10:00:17
4      2021-12-07 13:00:00
...
37417  2021-02-12 05:03:49
37418  2021-01-16 05:39:05
37419  2021-06-02 09:00:10
37420  2020-10-20 20:59:30
37421  2021-05-23 21:00:31
Name: published_at, Length: 37422, dtype: object
```

```
In [30]: 1 # -- Convert the published_at column from dtype-object to dtype-datetime64 using pandas datetime
2
3 df['published_at'] = pd.to_datetime(df['published_at'])
4 df['published_at']
```

```
Out[30]: 0      2021-07-01 10:00:00
1      2021-06-10 16:00:00
2      2021-09-20 01:03:32
3      2021-03-03 10:00:17
4      2021-12-07 13:00:00
...
37417  2021-02-12 05:03:49
37418  2021-01-16 05:39:05
37419  2021-06-02 09:00:10
37420  2020-10-20 20:59:30
37421  2021-05-23 21:00:31
Name: published_at, Length: 37422, dtype: datetime64[ns]
```

6. Create a new column as 'published_month' using the column published_at (display the months only)

```
In [31]: 1 # -- Created a new column called published_month and stored values of month in it.
          2
          3 df['published_month'] = df['published_at'].dt.month
          4 df['published_month']
```

```
Out[31]: 0      7
          1      6
          2      9
          3      3
          4     12
          ..
          37417    2
          37418    1
          37419    6
          37420   10
          37421    5
          Name: published_month, Length: 37422, dtype: int32
```

```
In [32]: 1 df
```

```
Out[32]:
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuW	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football alemn
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIellM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Qui Music/I
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE JO1 TheS
		Whv Retainin		Practical	2021-12-07					retainir

```
In [ ]: 1
```

7. Replace the numbers in the column published_month as names of the months i.e., 1 as 'Jan', 2 as 'Feb' and so on.....

```
In [33]: 1 # -- Converting numeric values to catagorical values using Dict
          2
          3 month_mapping = {1:'Jan', 2:'Feb', 3:'Mar', 4:'Apr', 5:'May', 6:'Jun', 7:'Jul', 8:'Aug', 9:'Sep', 10:'Oct', 11:'Nov', 12:'De
```

```
In [34]: 1 # -- Apply that converted values to published_month column using map() function
          2
          3 df['published_month'] = df['published_month'].map(month_mapping)
          4
```

```
In [35]: 1 df['published_month']
```

```
Out[35]: 0      Jul
          1      Jun
          2      Sep
          3      Mar
          4      Dec
          ...
          37417 Feb
          37418 Jan
          37419 Jun
          37420 Oct
          37421 May
          Name: published_month, Length: 37422, dtype: object
```

```
In [36]: 1 df
```

```
Out[36]:
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football socc alemn Bun se
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGleIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Av: Quality Music/Moto
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1*YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101 JO1 TheSTAF

8. Find the number of videos published each month and arrange the months in a decreasing order based on the video count.

```
In [37]: 1 # -- Count of the video_id based on the published_month
          2
          3 video_count_by_month = df.groupby(by='published_month')['video_id'].count()
          4 video_count_by_month
```

```
Out[37]: published_month
Apr      2126
Aug      4262
Dec      3072
Feb      2137
Jan      2108
Jul      2340
Jun      2316
Mar      2258
May      2081
Nov      4851
Oct      4991
Sep      4880
Name: video_id, dtype: int64
```

```
In [38]: 1 sorted_video_count = video_count_by_month.sort_values(ascending=False)
          2 sorted_video_count
```

```
Out[38]: published_month
Oct      4991
Sep      4880
Nov      4851
Aug      4262
Dec      3072
Jul      2340
Jun      2316
Mar      2258
Feb      2137
Apr      2126
Jan      2108
May      2081
Name: video_id, dtype: int64
```

9. Find the count of unique video_id, channel_id and channel_title.

```
In [39]: 1 #---Unique value counts of 'video_id' column
          2
          3 countof_unique_video_id = df['video_id'].nunique()
          4 countof_unique_video_id
```

Out[39]: 37422

```
In [40]: 1 #---Unique value counts of 'channel_id' column
          2
          3 countof_unique_channel_id = df['channel_id'].nunique()
          4 countof_unique_channel_id
```

Out[40]: 10961

```
In [41]: 1 #---Unique value counts of 'channel_title' column
          2
          3 countof_unique_channel_title = df['channel_title'].nunique()
          4 countof_unique_channel_title
```

Out[41]: 10883

10. Find the top10 channel names having the highest number of videos in the dataset and the bottom10 having lowest number of videos.

```
In [42]: 1 # -- count of the videos based on channel_title
          2
          3 channel_video_count = df.groupby('channel_title')['video_id'].count()
          4 channel_video_count
```

```
Out[42]: channel_title
          SilverName      1
          [KIMDUMB]       1
          (Inter TV channel) 1
          - IT!           1
          - MBN           1
          ..
          zanny           5
          zayn            1
          ze frank1       9
          zollotech       10
          zoom            1
          Name: video_id, Length: 10883, dtype: int64
```



```
In [43]: 1 top10_channel = channel_video_count.sort_values(ascending=False).head(10)
         2 top10_channel
```

```
Out[43]: channel_title
Sky Sports Football    533
The United Stand       301
BT Sport               246
NBA                   209
NFL                   162
WWE                   122
SSSniperWolf           99
SSundee                98
FORMULA 1              87
NHL                   86
Name: video_id, dtype: int64
```

```
In [44]: 1 bottom10_channel = channel_video_count.sort_values(ascending=True).head(10)
         2 bottom10_channel
```

```
Out[44]: channel_title
SilverName            1
Mini Muka             1
Mini Ladd             1
MindYourLogic         1
Mind Body Tonic With Dr Sita 1
Mimi Ar              1
Millyz               1
Milkair              1
Milissa Grande       1
MikuruSong           1
Name: video_id, dtype: int64
```

11. Find the title of the video which has the maximum number of likes and the title of the video having minimum likes and write your inferences.

In [45]:

1	df
---	----

Out[45]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football socc alemn Bun se
1	--14w5SOEU	Migos - Avalanche (Official Video)	UCGleIM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Ava Quality Music/Moto
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.) PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101 JO1 TheSTAF
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjIQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining w Jersey highwa C
...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Calling My Calling A
37418	zziBybeSATw	PELICANS at LAKERS FULL GAME HIGHLIGHTS Ja...	UCWJ2IWNubArHwMf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G Basketball 0022000187 L
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxlP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAMOO WAW MAM WAW Where
37420	zzmQEb0Em5I	FELLIPE ESCUERO- Master Podcast #12	UC8NjnNWMsRqq11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	masterpodc lord vinheteir
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	Spurs Tot Hotspur Tot Leic

37422 rows × 13 columns

```
In [46]: 1 # -- max likes based on title
2
3 title_max_likes = df.groupby('title')['likes'].max()
4 title_max_likes
```

```
Out[46]: title
          | Loki recipe          10675
          | /peanut          14594
          | # EP.38-01 #130529      10903
          | the Activist          12610
          | - Cornmeal Dosa & Chutney  22573
          | ...
you can pull the joystick up, but why?  333326
youtube striked lazarbeam :(          372547
z900rs          3581
zl grntlr: Azrbaycan sgrlri dedilr ki, bu df hrmt edirik, gedin  5877
zmir'de yaanan depremdede kedinin panik anlar  17297
Name: likes, Length: 37113, dtype: int64
```

```
In [47]: 1 max_likes = title_max_likes.sort_values(ascending=False).head(1)
2 max_likes
```

```
Out[47]: title
BTS () 'Dynamite' Official MV      31837675
Name: likes, dtype: int64
```

```
In [48]: 1 min_likes = title_max_likes.sort_values(ascending=True).head(1)
2 min_likes
```

```
Out[48]: title
Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News      0
Name: likes, dtype: int64
```

INFERENCE

```
In [49]: 1 '''
2 1) First, i groupby the title column and found the max likes using max() function on 'likes' column and stored that into one
3
4 2)Then, i sorted that max likes to descending order using sort_values(ascending=False) then i used head(1) to pull the first
5
6 3) Then, i sorted that min likes to ascending order using sort_values(ascending=True) then i used head(1) to pull the first
7
8 '''
9
```

12. Find the title of the video which has the maximum number of dislikes and the title of the video having minimum dislikes and write your inferences.

```
In [50]: 1 # -- max dislikes based on title
          2
          3 title_max_dislikes = df.groupby(['title'])['dislikes'].max()
          4 title_max_dislikes
```

```
Out[50]: title
          | Loki recipe                    1681
          | /peanut                      1677
          | # EP.38-01 #130529              793
          | the Activist                    815
          | - Cornmeal Dosa & Chutney          815
          | ...
you can pull the joystick up, but why? 12396
youtube striked lazarbeam :(           3289
z900rs                                 165
zl grntlr: Azrbaycan sgrlri dedilr ki, bu df hrmt edirik, gedin 168
zmir'de yaanan depremde kedinin panik anlar 1893
Name: dislikes, Length: 37113, dtype: int64
```

```
In [51]: 1 max_dislikes = title_max_dislikes.sort_values(ascending=False).head(1)
          2 max_dislikes
```

```
Out[51]: title
Cuties | Official Trailer | Netflix    2397733
Name: dislikes, dtype: int64
```

```
In [52]: 1 min_dislikes = title_max_dislikes.sort_values(ascending=True).head(1)
          2 min_dislikes
```

```
Out[52]: title
Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News    0
Name: dislikes, dtype: int64
```

Inference

```
In [53]: 1 '''
          2 1) First, i groupby the title column and found the max dislikes using max() function on 'dislikes' column and stored that in
          3
          4 2)Then, i sorted that max dislikes to descending order using sort_values(ascending=False) then i used head(1) to pull the fi
          5
          6 3) Then, i sorted that min dislikes to ascending order using sort_values(ascending=True) then i used head(1) to pull the fir
          7
          8 '''
```

13. Does the number of views have any effect on how many people disliked the video? Support your answer with a metric and a plot

```
In [54]: 1 correlation = df['view_count'].corr(df['dislikes'])  
        2 correlation
```

Out[54]: 0.6844687753905547

```
In [55]: 1 print("Correlation between views and dislikes:", correlation)
```

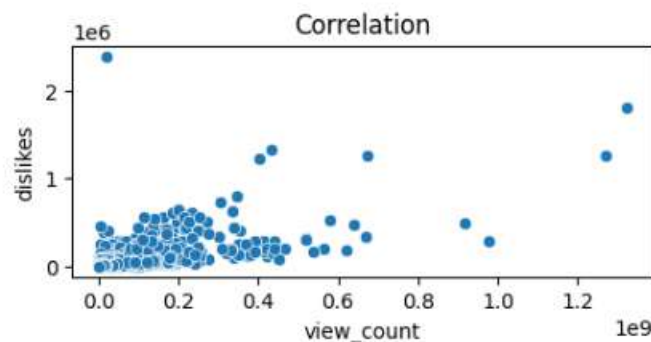
Correlation between views and dislikes: 0.6844687753905547

```
In [56]: 1 # Analyze the significance of the correlation  
        2  
        3 if correlation > 0:  
        4     print("There is a positive correlation between views and dislikes.")  
        5 elif correlation < 0:  
        6     print("There is a negative correlation between views and dislikes.")  
        7 else:  
        8     print("There is no linear correlation between views and dislikes.")
```

There is a positive correlation between views and dislikes.

Plot

```
In [57]: 1 import matplotlib.pyplot as plt  
        2 import seaborn as sns  
        3  
        4 plt.figure(figsize=(5,2))  
        5 sns.scatterplot(data=df,x = 'view_count', y = 'dislikes')  
        6 plt.title(label='Correlation')  
        7 plt.show()
```



14. Display all the information about the videos that were published in January, and mention the count of videos that were published in January.

In [61]:

```
1 # -- All the information about the video that were published in january
2
3 info_of_videos = df[df['published_month'] == 'Jan']
4 info_of_videos
```

Out[61]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tags
27	-2Gwm7QfBnE	Q&A With Naisha	UCYwNMbogQFzMccPSuy-pPWg	MianTwins	2021-01-21 00:05:47	872372	38626	239	621	He
48	-4sfXSHSxzA	SURPRISING BRENT WITH HIS TIKTOK CRUSH!!	UCPpATKqmMV-CNRNWYyADUwiA	Alexa Rivera	2021-01-16 21:40:04	6504784	262477	5779	7907	
95	-AJD1Fc5rpQ	WE ARE HAVING A BABY! finding out i'm pregna...	UCVsTboAhpnuL6j-tDePvNwQ	Tess Christine	2021-01-03 21:53:48	533084	38965	119	1650	I a
103	-AuJiwjsmWk	Do Ugly Foods Taste Worse? Taste Test	UCZpCc5n9hqiVC7HhPwclKEg	Good Mythical MORE	2021-01-19 11:00:01	1057077	22526	531	773	gmm good mythical morning rhettandlink rhett a...
182	-JhqO2KWw5U	Schlatt gets fit	UCWZp4y1jqBuvLtiyxSs_ZBw	Big guy	2021-01-24 22:50:57	1724965	119431	325	1578	jschlatt big guy jschlatt highlights schlatt j...
...
37300	zmzFL5bG-jc	DEVINE MON PERSONNAGE AVANT AKINATOR ! (c'est ...	UCllr3byh6wmXgcPx_Tm9Ocw	Piwerre	2021-01-16 16:12:19	670357	54462	832	1249	Piwerre frere de michou crouton among us devin... http:
37329	zpzjex7qwrA	Lampard Sacked Within Days Rorys Misery Chel...	UCkD-ZOixl0a9FjlExDsHsbg	The Kick Off	2021-01-03 20:13:49	428646	12060	296	1505	Premier league Chelsea chelsea 1-3 Man City Ch... N
37345	zqyv-B6mnBM	Lil Wayne - Ain't Got Time (Audio)	UCO9zJy7HWrlS3ojB4Lr7Yqw	Lil Wayne	2021-01-21 05:00:10	2238244	58925	2365	5539	lil wayne weezy weezy wednesday

37329	zpzjex7qwrA	Lampard Sacked Within Days Rorys Misery Chel...	UCkD-ZOixl0a9FjlExDsHsbg	The Kick Off	2021-01-03 20:13:49	428646	12060	296	1505	Premier league Chelsea 1-3 Man City Ch...	N
37345	zqyv-B6mnBM	Lil Wayne - Ain't Got Time (Audio)	UCO9zJy7HWrlS3ojB4Lr7Yqw	Lil Wayne	2021-01-21 05:00:10	2238244	58925	2365	5539	lil wayne weezy weezy wednesday wayne carter y...	
37383	zwfu1-24T7Q	PRADA Cup Day 1 Full Race Replay PRADA Cup...	UCo15ZYO_XDRU9LI30OPTxAg	America's Cup	2021-01-15 04:07:55	317382	2008	83	192	America's Cup Americas Cup AC36 AC75 Presented...	F
37418	zziBybeSAtw	PELICANS at LAKERS FULL GAME HIGHLIGHTS Ja...	UCWJ2IWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G League Basketball game-0022000187 Lakers...	PEI

2108 rows × 13 columns



```
In [63]: 1 df[df['published_month'] == 'Jan']['video_id'].count()  #-- Count of videos that were published in january
```

Out[63]: 2108

```
In [ ]: 1
```